Arabiyatuna
Jurnal Bahasa Arab

Vol. 9. No. 2, 2025,
P-ISSN: **2580-5045, E-ISSN: 2580-5053**
DOI: 10.29240/jba.v9i2.13889

Development of HOTS-Based Final Semester Assessment Questions in the *Al-'Ashri* Arabic Textbook for 10th-grade Students of Islamic Senior High School

Khusnul Masrifah[1], Mohammad Ahsanuddin[2]
Universitas Negeri Malang[1,2]
khusnul.masrifah.2302318@students.um.ac.id[1],
mohammad.ahsanuddin.fs@um.ac.id[2]

Cite this article:
Masrifah, Khusnul and Ahsanuddin, Mohammad. (2025). Development of HOTS-Based Final Semester Assessment Questions in the *Al-'Ashri* Arabic Textbook for 10th-grade Students of Islamic Senior High School. Arabiyatuna: Jurnal Bahasa Arab, 9(2), 754-769. doi: 10.29240/jba.v9i1.13889

Received: 18-06-2025	Revised: 07-10-2025	Accepted: 14-10-2025

Abstract

This study aimed to develop and validate a Higher-Order Thinking Skills (HOTS)–based assessment instrument for Arabic language learning using the *Al-'Ashri* textbook for 10th-grade students at an Islamic Senior High School. The research employed a Research and Development (R&D) design following the 4D model: Define, Design, Develop, and Disseminate. Several research instruments were utilized, including: (1) a needs analysis questionnaire and interview guidelines to identify issues in existing Arabic learning assessments; (2) validation sheets for material experts and linguists to evaluate content, construct, and language validity; (3) a HOTS-based test instrument consisting of 36 multiple-choice and 4 essay questions developed according to Bloom's revised taxonomy (C4–C6); and (4) data analysis tools to measure validity, reliability, item difficulty, and discrimination indices processed using SPSS. Each instrument served a distinct purpose: the questionnaire and interviews gathered empirical data on field needs, the validation sheets assessed the feasibility and accuracy of content and language, and the HOTS-based test was empirically tested on 34 students to examine its psychometric properties. The material validation achieved a score of 94% and the language validation 84%, both categorized as "highly feasible." Empirical testing showed that all items were valid (r-value > r-table = 0.338), with a reliability coefficient of r = 0.88 (high), item

difficulty indices ranging from 0.15–0.85 (moderate), and discrimination indices indicating 38.89% good and 61.11% moderate items. These findings confirmed that the developed HOTS-based assessment instrument was valid, reliable, and effective for evaluating higher-order thinking skills in Arabic language learning at the senior high school level.

Keywords: Arabic Textbook; Final Semester Assessment; Higher-Order Thinking Skills.

Introduction

Educational assessment aims not only to measure learning outcomes but also to improve the learning process. To conduct effective assessments, teachers require well-designed instruments in the form of questions that can measure students' cognitive, affective, and psychomotor abilities. Therefore, such questions need to be analyzed and developed based on Higher-Order Thinking Skills (HOTS), referring to Bloom's Taxonomy.[1] Learning outcomes are statements that describe the achievements of learning, which may include the acquisition of knowledge, comprehension, and intellectual or practical skills.[2] HOTS involves the ability to analyze, evaluate, and create, beyond simple recall or understanding. The essence of HOTS lies in students' ability to solve problems, ask questions, reason, communicate, and connect concepts.[3] In implementing HOTS in learn

ing, teaching materials or textbooks are crucial components to consider in lesson planning, delivery, and the development of students' higher-order

---

[1] Noza Aflisia, " تطوير المواد التعليمية النحوية على أساس معرفة المحتوى التربوي التكنولوجي لتنمية مهارات التفكير العليا - لدى طلاب قسم تعليم اللغة العربية في جامعة جوروب الإسلامية الحكومية وجامعة فتماواتي سوكرنو الإسلامية الحكومية بنجكولو Digital Library UIN Sunan Gunung D" (2023), https://digilib.uinsgd.ac.id/71137/.

[2] P N Sagala and A Andriani, "Development of Higher-Order Thinking Skills (HOTS) Questions of Probability Theory Subject Based on Bloom's Taxonomy," *Journal of Physics: Conference Series* 1188 (March 2019): 012025, https://doi.org/10.1088/1742-6596/1188/1/012025; Johan Setiawan et al., "Development of Higher Order Thinking Skill Assessment Instruments in Learning Indonesian History," *International Journal of Evaluation and Research in Education (IJERE)* 10, no. 2 (June 2021): 545, https://doi.org/10.11591/ijere.v10i2.20796.

[3] Tika Mardiyah, Umi Machmudah, and Abdul Wahab Rosyidi, "Faáliyah Al-Tadribat 'ala Asas Maharat Al-Tafkir Al-Úlya Fi Ta'lim Al-Qiraáh Al-Árabiyyah," *Arabiyatuna : Jurnal Bahasa Arab* 5, no. 2 (October 25, 2021): 289, https://doi.org/10.29240/jba.v5i2.3064; Bansu Irianto Ansari et al., "Exploring Students' Learning Strategies and Self-Regulated Learning in Solving Mathematical Higher-Order Thinking Problems," *European Journal of Educational Research* volume-10-, no. volume-10-issue-2-april-2021 (April 2021): 743–56, https://doi.org/10.12973/eu-jer.10.2.743; Aleksius Madu, "Higher Order Tingking Skills (Hots) In Math Learning," *IOSR Journal of Mathematics (IOSR-JM)* 13, no. 5 (2017): 70–75, https://doi.org/https://doi.org/10.9790/5728-1305027075.

thinking.[4] Textbooks are more than their physical appearance or cover; the way the material is arranged and presented inside also plays an important role in attracting readers and helping them understand the content more easily.[5]

In Arabic language learning, most practice questions remain at the Lower-Order Thinking Skills (LOTS) and Middle-Order Thinking Skills (MOTS) levels. Consequently, the lack of HOTS-based assessments results in limited student engagement and suboptimal learning outcomes. One of the widely used Arabic textbooks in East Java is the *Al-'Ashri* Arabic textbook, particularly in Muhammadiyah schools. This textbook has been proven to meet the National Education Standards (BSNP) in terms of content, presentation, and language quality.[6] Studies have confirmed that it provides coherent material organization, a communicative language style, and strong pedagogical alignment with curriculum objectives. Similarly, other research has shown that the same textbook effectively supports students' speaking skills through contextual materials consistent with the Common European Framework of Reference for Languages (CEFR).[7] Furthermore, studies indicate that Arabic textbooks adhering to BSNP standards and integrated with HOTS-based principles enhance students' analytical and interpretative abilities, thereby fostering more meaningful language learning experiences.[8,9] These consistent findings suggest that the *Al-'Ashri* textbook not only fulfills national textbook quality benchmarks but also aligns with contemporary learning paradigms that emphasize critical, creative, and communicative competence.

Therefore, this study aimed to develop and validate HOTS-based final semester assessment questions for the *Al-'Ashri* Arabic textbook used by 10th-grade students at an Islamic senior high school. The developed instrument is expected to enhance students' analytical, evaluative, and creative thinking in

---

[4] Sutri Ramah and Miftahur Rohman, "Analisis Buku Ajar Bahasa Arab Madrasah Aliyah Kurikulum 2013," *Arabiyatuna : Jurnal Bahasa Arab* 2, no. 2 (2018): 141, https://doi.org/10.29240/jba.v2i2.552; Maratul Qiftiyah, "Muatan HOTS Pada Pembelajaran Tematik Materi IPA Kelas 5 Sekolah Dasar," *Scholaria: Jurnal Pendidikan Dan Kebudayaan* 13, no. 1 (January 2023): 28–38, https://doi.org/10.24246/j.js.2023.v13.i1.p28-38.

[5] Noza Aflisia, Rizka Salsabillah, and Asma Binti Abdul Rahman, "Review of Arabic Language Books Independent Curriculum in Madrasah Ibtidaiyah: Materials, Methods, Advantages, and Disadvantages," *Al-Muktamar As-Sanawi Li Al-Lughah Al-'Arabiyyah (MUSLA)* 2, no. 1 (October 9, 2024): 85–100, https://doi.org/10.24042/albayan.v14i2.12514.

[6] Khizanatul Hikmah and Ruli Astuti, "Analisis Perbandingan Kualitas Buku Teks Bahasa Arab Ta'lim Al-Lughoh Al-'Arobiyah Dan Al-'Ashri: Kajian Isi, Penyajian, Dan Bahasa," *Halaqa: Islamic Education Journal* 2, no. 1 (2018): 12–29, https://doi.org/10.21070/halaqa.v1i1.1608.

[7] Salim Saputra and Tulus Musthofa, "Analisis Buku Teks Pelajaran Bahasa Arab SMA/SMK Muhammadiyah Pada Maharah Al-Kalam Sesuai Standar CEFR," *Lugawiyyat* 4, no. 2 (2022): 81–97, https://doi.org/10.18860/lg.v4i2.17393.

[8] Ramah and Rohman, "Analisis Buku Ajar Bahasa Arab Madrasah Aliyah Kurikulum 2013."

[9] Arif Widodo et al., "Development of Arabic Learning Design Based on Higher Order Thinking Skill with Discovery Learning Model," *Al Mahāra: Jurnal Pendidikan Bahasa Arab* 7, no. 1 (2021): 27–44, https://doi.org/10.14421/almahara.2021.071-02.

Arabic learning. This research employed a Research and Development (R&D) approach[10] using the 4D model (Define, Design, Develop, and Disseminate) proposed by Thiagarajan (1974). This model was chosen because it is designed to produce a product; in this case, a HOTS-based assessment instrument capable of measuring students' higher-order thinking skills in Arabic learning. To ensure that the developed product is valid, reliable, and applicable, several research instruments were used. The needs analysis instruments included structured observation sheets, teacher and student questionnaires, and interview guidelines to collect data on current Arabic learning and assessment practices. These instruments used a Likert-scale rating (1–5), resulting in expert validation scores of 94% for content and 84% for language, both categorized as highly feasible. The test instrument consisted of 36 multiple-choice and 4 essay questions, developed based on Bloom's revised taxonomy (C4–C6). For empirical testing, data analysis included item validity tests (using Pearson's product-moment correlation), reliability tests (split-half and Spearman–Brown formula), difficulty index, and discrimination index analyses, all processed through SPSS.

The development of HOTS-based question exercises as assessment instruments aims to train the higher-order thinking skills of 10th-grade students. This effort contributes to the enrichment of Arabic textbook evaluation instruments and serves as a reference for future assessments, helping to improve students' critical thinking skills as foreign language learners.

Findings and Discussion

Define Stage

At this stage, various necessary information to establish the preparation and requirements for product development were collected. The following are the results of the five analysis activities that were carried out.

First, Initial Analysis

At this stage, identification of research problems through needs analysis was conducted using observations and interviews with Arabic language teachers at SMA Muhammadiyah 1 Ponorogo.

Observations showed that the learning process followed the *Merdeka* Curriculum. The teacher implemented the direct method, where learning took place with the teacher presenting the material orally, clearly pronouncing each word with the aid of pictures, and students repeating it several times until they pronounced it correctly and understood its meaning. This activity was interspersed with question-and-answer sessions, followed by reading the text

[10] Ramadhan Jabal Primadana et al., "Development of Listening and Speaking Skills Teaching Materials Based on CEFR Level," *Arabiyatuna: Jurnal Bahasa Arab* 8, no. 2 (September 30, 2024): 559–78, https://doi.org/10.29240/jba.v8i2.10904.

alternately and discussing it to ensure comprehension. The material was presented with the help of PowerPoint, and sometimes smartphones were used to find the meaning of new vocabulary for discussion. After the lesson, the teacher gave exercises and used the student handbook as a learning resource. The book used was the *Al-'Ashri* Arabic textbook published by the DIKDASMEN Muhammadiyah. Initial observations also indicate that students were given opportunities to discuss the content of reading texts on materials that supported higher-order thinking.

Furthermore, interviews with Arabic language teachers revealed that the type of assessment instrument used was formative assessment, conducted continuously during the learning process, such as through quizzes, discussions, and direct question-and-answer sessions. Teachers acknowledged that the level of HOTS among students in Arabic learning was still not optimal. They had not yet used assessment instruments aimed at improving HOTS due to time constraints. Therefore, a summative assessment instrument was needed to train students' HOTS and support more effective learning.

Second, Learner Analysis

The analysis revealed that 10th-grade senior high school students (aged 15–16) generally possessed basic conceptual understanding but required further guidance in applying theory to practice. Most students exhibited visual and kinesthetic learning preferences, responding better to interactive and hands-on activities than to passive instruction. Their motivation was relatively high when learning contextual topics relevant to daily life, though interest tended to decline when dealing with abstract grammatical concepts. Cognitively, students were still at the C1–C2 levels of Bloom's taxonomy, often struggling to apply knowledge in analytical or problem-solving contexts. This finding indicates the need for instructional materials that strengthen HOTS.

***Third,* Task Analysis**

The task analysis showed that the existing questions in the *Al-'Ashri* Arabic textbook largely remained at the LOTS level, focusing on recall and comprehension rather than analysis, evaluation, or creation. The absence of summative assessments measuring higher-order cognitive abilities indicates the need to develop HOTS-based instruments to complement the textbook and enhance students' analytical and creative thinking in Arabic learning.

Fourth, Concept Analysis

The purpose of this analysis was to identify, elaborate, and systematically arrange the main concepts of the material to be developed. The analysis of Arabic language materials for 10th-grade students in the first semester was carried out based on the facts, concepts, principles, and theories contained therein. The ideas derived from this analysis served as the foundation for developing assessment instruments to foster HOTS.

Fifth, Formulation of Learning Objectives

In analyzing learning objectives, each learning activity was aligned with learning outcomes, learning progression, and corresponding learning indicators. Based on this, the grid of questions to be developed could be determined. The results of this analysis are presented in Appendix 1e. Based on these studies, it can be concluded that students needed more opportunities to train their higher-order thinking skills. Therefore, the researchers decided to develop the end-of-semester exercises.

Design Stage

Based on the results of the analysis conducted in the first stage, this stage involved the creation of Draft 1 as the initial product design. The following four activities were carried out.

First, Compilation of Question Grids

The test grid aimed to facilitate the process of developing questions. With the grid, the questions became more focused and aligned with the researchers' objectives. The question grid was developed based on HOTS indicators. The following was the question grid for the HOTS-based assessment instrument on Arabic language material found in the *Al-'Ashri* textbook, as presented in Table 1 below.

Table 1. HOTS-Based Test Guidelines

Theme	Skill	Question Indicator	No	Question Format	Cognitive Level
التَّعَارُفُ	*Qira'ah*	Finding the main idea	1	Multiple Choice	C4 (Analyzing)
		Finding the message from the text provided	2		
		Finding the reading title	3		
		Identifying relevant information related to the text *At-Ta'aruf*	4		
		Commenting on the author's ideas related to the text *At-Ta'aruf*	5	Multiple Choice	C5 (Evaluating)
		Assessing the author's feelings/impressions in the *At-Ta'aruf* text	6		
	Qawaid	Identifying sentence structures in accordance with the rules of *nahwiyah dhomir munfasil*	7,8,9	Multiple Choice	C4 (Analyzing)
	Kitabah	Writing a short story related to *At-Ta'aruf*	1	Essay	C6 (Creating)
البَيْتُ	*Qira'ah*	Finding the message from the text provided	10	Multiple Choice	C4 (Analyzing)
		Identifying relevant information related to *Al-Baitu* text	11, 15		

		Finding the main idea	12		
		Assessing the author's feelings/impressions in the *Al-Baitu* text	13	Multiple Choice	C5 (Evaluating)
		Commenting on the author's ideas related to the *Al-Baitu* text	14		
		Expressing opinions related to keeping the house clean	16		
	Qawaid	Identifying sentence structures in accordance with the rules of *nahwiyah dhomir muttasil*	17	Multiple Choice	C4 (Analyzing)
		Comparing differences in sentence structure	18		
	Kitabah	Narrating a sequence of pictures related to daily activities on the theme *Al-Baitu*	3	Essay	C6 (Creating)
	Qira'ah	Identifying relevant information related to the text *Al-Madrasatu*	19,22,26	Multiple Choice	C4 (Analyzing)
الْمَدْرَسَةُ		Finding the message from the text provided	20		
		Finding the main idea	21		
		Finding the reading title	23		
		Commenting on the author's ideas related to the text *At-Ta'aruf*	24	Multiple Choice	C5 (Evaluating)
		Assessing the author's feelings/impressions in the *At-Ta'aruf* text	25		
	Qawaid	Identifying sentence structures in accordance with the rules of *nahwiyah* on the chapter *jama'*	27	Multiple Choice	C4 (Analyzing)
	Kitabah	Composing sentences based on pictures related to school activities	4	Essay	C6 (Creating)
	Qira'ah	Finding the message from the text provided	28	Multiple Choice	C4 (Analyzing)
		Finding the main idea	29		
الْمَكْتَبَةُ		Identifying relevant information related to the text *Al-Maktabatu*	30		
		Finding the reading title	31		

	Assessing the author's feelings/impressions in the *Al-Maktabatu* text	32	Multiple Choice	C5 (Evaluating)
	Commenting on the author's ideas related to the *Al-Maktabatu* text	33		
Qawaid	Identifying sentence structures in accordance with the rules of *nahwiyah harfu jar*	34,35,36	Multiple Choice	C4 (Analyzing)
Kitabah	Writing a short story related to the theme *Al-Maktabatu* and *harfu jar*	2	Essay	C6 (Creating)

Second, Media Selection

The learning evaluation media used in this study referred to paper-based assessment media. This study employed conventional paper-based test materials, using printed question sheets and answer sheets commonly used in school examinations.

Third, Initial Design

The initial design (Draft 1) consisted of the development of assessment instruments and answer sheets. The questions were designed to measure critical and higher-order thinking skills. The assessment instrument included 36 multiple-choice questions and 4 essay questions, all structured according to HOTS indicators. The instrument was administered using printed sheets similar to regular daily or semester tests, and students recorded their answers on provided answer sheets. An answer key was also prepared, accompanied by clear guidelines for scoring and instructions for responding to each question.

Development Stage (Develop)

The output of this stage was the final product of the HOTS assessment instrument, which was developed and revised to be ready for dissemination.

First, Material Expert Validation

Table 2. Material Expert Validation Results

No	Assessment Aspect	Score (%)	Category
1	Relevance of test items to learning objectives	96	Highly feasible
2	Alignment with core and basic competencies	95	Highly feasible
3	Conformity with HOTS indicators (C4–C6)	93	Highly feasible
4	Accuracy and clarity of question content	92	Highly feasible
5	Logical consistency of answer options	94	Highly feasible
6	Completeness of supporting components (grid, key, scoring)	94	Highly feasible
Average Score		94%	Highly feasible

Material validation was conducted by expert lecturers to determine the validity of the developed product. The validation process employed a structured questionnaire assessing the suitability of the material with learning outcomes, the logical flow of learning objectives, and its alignment with HOTS indicators. Material experts provided evaluations, suggestions, and comments on the HOTS-based Arabic assessment instrument for the 10th-grade students in their first semester.

Recommendations included clarifying the number of paragraphs in writing tasks, arranging essay questions from easy to difficult, limiting writing tasks to *kitabah qissah qosirah*, and providing clear scoring criteria. After implementing these revisions, the instrument achieved a 94% validation score, categorized as highly feasible for testing. The developed test items aligned well with the learning materials, core and basic competencies, learning objectives, and Bloom's taxonomy levels. The items were relevant to HOTS indicators, logically structured, free from multiple interpretations, and supported by complete assessment components (grids, items, answer sheets, and scoring rubrics).

Second, Linguist Validation

Table 3. Linguist Validation Results

No	Assessment Aspect	Score (%)	Category
1	Grammatical accuracy (*qawā'id*)	82	Feasible
2	Accuracy of *harakat* and spelling	83	Feasible
3	Word choice and diction appropriateness	85	Highly feasible
4	Readability and clarity of sentences	84	Highly feasible
5	Consistency in Arabic and Indonesian terms	86	Highly feasible
Average Score		84%	Highly feasible

The linguistic validation aimed to assess the accuracy and clarity of the language used in the assessment instrument, including sentence structure, grammar (*qawā'id*), spelling, diction, and paragraph cohesion. Validation was conducted using a linguist validation questionnaire. Suggestions from linguistic experts included: (1) Re-examining and correcting *harakat* placement; (2) Reviewing the use of *ma'rifat* and *nakirah;* and (3) Revising sentence structures and wording according to corrections.

All inputs from the linguistic experts were incorporated into the revision process. As a result, the linguistic quality of the instrument met academic writing standards. The final validation score was 84%, categorized as highly feasible.

The question items were written in accordance with Arabic grammatical rules, used easily understandable language, maintained readability and clarity, adhered to spelling and punctuation conventions, and ensured consistency between Arabic and Indonesian terms.

Third, Product Trial

The product trial stage aimed to identify the feasibility level and potential weaknesses of the instrument, including its effectiveness and readability by the

target group. The objective was to ensure the instrument met evaluation standards and eliminated errors in the narrative or structure. Data were collected through the trial implementation of the HOTS-based Arabic assessment instrument for 10th-grade students on 34 students from Class X.1 and Class X.2.

In the first product trial, the validity test using SPSS on 36 questions showed that only 18 items were valid. This indicated that the instrument had not yet met the required validity and discrimination indices, thus requiring further revision. In the second product trial, after revising the items, the HOTS-based Arabic assessment instrument met the criteria for validity, reliability, difficulty level, and discrimination power, and was therefore deemed suitable for use.

Fourth, Validity Test

The item validity was calculated using the SPSS program. The number of 10th-grade students who participated as test subjects was 34. The results of the calculation were then compared to the r-table at a significance level of 0.05 (5%)[11]. The criteria used in the validity analysis were as follows: if r-value ≥ r-table, then the item was considered valid; conversely, if r-value < r-table, then the item was declared invalid. The results of the validity test for the development of Arabic HOTS question exercises for 10th-grade students are presented in the following table.

Table 4. Final Results of the Validity Test

No	r count	r table	Description	No	r count	r table	Description
1	0,602**	0,338	valid	19	0,394*	0,338	valid
2	0,365*	0,338	valid	20	0,603**	0,338	valid
3	0,445**	0,338	valid	21	0,359*	0,338	valid
4	0,626**	0,338	valid	22	0,527**	0,338	valid
5	0,443**	0,338	valid	23	0,345*	0,338	valid
6	0,450**	0,338	valid	24	0,390*	0,338	valid
7	0,402*	0,338	valid	25	0,343*	0,338	valid
8	0,376*	0,338	valid	26	0,525**	0,338	valid
9	0,427*	0,338	valid	27	0,449**	0,338	valid
10	0,434*	0,338	valid	28	0,475**	0,338	valid
11	0,591**	0,338	valid	29	0,398*	0,338	valid
12	0,492**	0,338	valid	30	0,443**	0,338	valid
13	0,393*	0,338	valid	31	0,405*	0,338	valid
14	0,368*	0,338	valid	32	0,359*	0,338	valid
15	0,434*	0,338	valid	33	0,372*	0,338	valid
16	0,394*	0,338	valid	34	0,367*	0,338	valid
17	0,385*	0,338	valid	35	0,429*	0,338	valid
18	0,468**	0,338	valid	36	0,394*	0,338	valid

[11] Miske Hayunia Hamidah and Siti Sri Wulandari, "Pengembangan Instrumen Penilaian Berbasis HOTS Menggunakan Aplikasi 'QUIZIZZ,'" *Efisiensi : Kajian Ilmu Administrasi* 18, no. 1 (May 2021): 105–24, https://doi.org/10.21831/efisiensi.v18i1.36997.

A total of 34 10th-grade students worked on the HOTS-based assessment instrument, which consisted of 36 items, with two correlated variables. To determine which items were valid or invalid, the value of the r-table must first be identified. The formula used to find the r-table was df = N – 2, where N represents the number of students. Thus, df = 34 – 2 = 32. Referring to the distribution table for the r-table at a 0.05 (5%) significance level, the obtained r-table for N = 32 was 0.338. Therefore, an item could be considered valid if r-value ≥ 0.338. Based on the results of the validity test for the 36 multiple-choice items in the HOTS-based Arabic exercise for 10th-grade students, it was found that all r-values were greater than the r-table. Hence, all items were declared valid.

Fifth, Reliability Test

The reliability test analysis was carried out using the split-half technique, namely by dividing the scores into two groups: the odd group (X) and the even group (Y). These were then correlated using the Product-Moment formula to obtain the reliability level for half of the test items. Subsequently, the Spearman-Brown formula was applied to determine the reliability coefficient for the entire set of items. If the correlation coefficient ranges between 0.90–1.00, the reliability is categorized as very high; between 0.70–0.89 as high; between 0.50–0.69 as moderate; between 0.30–0.49 as low; and below 0.30 as very low.[12] The odd and even score data are presented in the following table.

Table 5. Odd and Even Score Results

No	X	Y	X^2	Y^2	XY
1	13	10	169	100	130
2	16	12	256	144	192
3	14	17	196	289	238
4	9	14	81	196	126
5	17	15	289	225	255
6	7	11	49	121	77
7	11	11	121	121	121
8	14	12	196	144	168
9	12	12	144	144	144
10	15	14	225	196	210
11	14	14	196	196	196
12	12	13	144	169	156
13	12	12	144	144	144
14	13	15	169	225	195
15	15	13	225	169	195
16	12	13	144	169	156
17	7	9	49	81	63
18	15	12	225	144	180
19	13	12	169	144	156
20	11	8	121	64	88
21	8	8	64	64	64
22	9	11	81	121	99
23	11	11	121	121	121
24	9	11	81	121	99
25	6	3	36	9	18
26	9	9	81	81	81
27	14	15	196	225	210
28	5	7	25	49	35
29	16	15	256	225	240
30	6	8	36	64	48
31	5	2	25	4	10
32	15	12	225	144	180
33	3	3	9	9	9
34	8	2	64	4	16
∑	376	366	4612	4426	4420

[12] Moh Ainin, *Penilaian Berpikir Tingkat Tinggi (HOTS) Dalam Pembelajaran Bahasa Arab* (Malang: CV. Bintang Sejahtera, 2020).

Based on the results of the reliability test for the question items, the obtained value was 0.88. This result is supported by the correlation index, which indicates that all items have a high level of reliability. Therefore, it can be concluded that the test instrument has good internal consistency and is suitable for use in evaluative research or measurement.

Sixth, Difficulty Level

The item difficulty level represents the proportion of students who answered each question correctly relative to the total number of test participants.

Table 6. Analysis of Item Difficulty Level

Question	∑ Correct Answer	∑ Test takers	Difficulty Result	Description
Q1	22	34	0.65	Medium (Feasible)
Q2	10	34	0.29	Medium (Feasible)
Q3	22	34	0.65	Medium (Feasible)
Q4	26	34	0.76	Medium (Feasible)
Q5	15	34	0.44	Medium (Feasible)
Q6	29	34	0.85	Medium (Feasible)
Q7	29	34	0.85	Medium (Feasible)
Q8	15	34	0.44	Medium (Feasible)
Q9	28	34	0.82	Medium (Feasible)
Q10	16	34	0.47	Medium (Feasible)
Q11	25	34	0.74	Medium (Feasible)
Q12	23	34	0.68	Medium (Feasible)
Q13	21	34	0.62	Medium (Feasible)
Q14	17	34	0.50	Medium (Feasible)
Q15	16	34	0.47	Medium (Feasible)
Q16	28	34	0.82	Medium (Feasible)
Q17	23	34	0.68	Medium (Feasible)
Q18	25	34	0.74	Medium (Feasible)
Q19	28	34	0.82	Medium (Feasible)
Q20	29	34	0.85	Medium (Feasible)
Q21	17	34	0.50	Medium (Feasible)
Q22	15	34	0.44	Medium (Feasible)
Q23	18	34	0.53	Medium (Feasible)
Q24	19	34	0.56	Medium (Feasible)
Q25	17	34	0.50	Medium (Feasible)
Q26	27	34	0.79	Medium (Feasible)
Q27	26	34	0.76	Medium (Feasible)
Q28	13	34	0.38	Medium (Feasible)
Q29	8	34	0.24	Medium (Feasible)
Q30	15	34	0.44	Medium (Feasible)
Q31	28	34	0.82	Medium (Feasible)
Q32	27	34	0.79	Medium (Feasible)
Q33	13	34	0.38	Medium (Feasible)
Q34	23	34	0.68	Medium (Feasible)
Q35	20	34	0.59	Medium (Feasible)
Q36	9	34	0.26	Medium (Feasible)

Based on the table above, the difficulty level analysis conducted on the 36 items produced values ranging from 0.15 to 0.85. This indicates that, overall, the items fall into the medium category. Thus, the HOTS-based question exercise instrument is considered appropriate in terms of item difficulty level and suitable for use.

Seventh, Discriminating Power

The discriminating power analysis aimed to differentiate between students who had mastered the competencies and those who had not. The upper and lower

groups were determined by taking 27% of the total test participants, calculated as 27% × 34 = 9.18. The results of the discriminating power analysis are presented in the following table.

Table 7. Analysis of Question Item Discriminating Power

∑ True Upper			∑ True Lower	Uper/Lower Group	Differentiating Power Result				Description
8	1	9	0.78	Good	8	5	9	0.33	Medium
5	1	9	0.44	Medium	9	5	9	0.44	Medium
7	3	9	0.44	Medium	6	3	9	0.33	Medium
9	3	9	0.67	Good	7	0	9	0.78	Good
7	1	9	0.67	Good	7	3	9	0.44	Medium
8	6	9	0.22	Medium	7	3	9	0.44	Medium
9	6	9	0.33	Medium	6	3	9	0.33	Medium
7	1	9	0.67	Good	8	4	9	0.44	Medium
9	6	9	0.33	Medium	9	5	9	0.44	Medium
6	1	9	0.56	Good	6	0	9	0.67	Good
9	3	9	0.67	Good	4	0	9	0.44	Medium
9	4	9	0.56	Good	7	1	9	0.67	Good
9	4	9	0.56	Good	8	5	9	0.33	Medium
5	3	9	0.22	Medium	8	5	9	0.33	Medium
7	1	9	0.67	Good	7	2	9	0.56	Good
9	6	9	0.33	Medium	8	4	9	0.44	Medium
7	3	9	0.44	Medium	7	3	9	0.44	Medium
8	3	9	0.56	Good	4	1	9	0.33	Medium

Based on the table above, the analysis results show that, out of the 36 items tested, 14 items (38.89%) were categorized as good, and 22 items (61.11%) were categorized as moderate. Therefore, all items are considered feasible and suitable for use in the high-order thinking assessment instrument.

Disseminate Stage

In the final stage, product dissemination, the researchers produced a final product in the form of a HOTS-based assessment instrument for Arabic language material for 10th-grade senior high school students in the first semester. This instrument had been revised and met the criteria for good test characteristics, making it suitable for distribution and use. In this stage, dissemination was limited to the school that served as the research site and was conducted as part of the research focus. The instrument was provided to the Arabic language teacher at SMA Muhammadiyah 1 Ponorogo, in both printed (hard copy) and digital (soft copy) formats. The materials included a question grid, HOTS-based Arabic question exercises, and accompanying guidelines and scoring rubrics. Examples of the HOTS-based Arabic test items are presented in the appendix below:

اِسْمِي أَحْمَدُ. أَنَا تِلْمِيذٌ فِي الْفَصْلِ الْعَاشِرِ فِي الْمَدْرَسَةِ الثَّانَوِيَّةِ الْمُحَمَّدِيَّةِ. كُنْتُ مُتَخَرِّجًا فِي الْمَدْرَسَةِ الْمُتَوَسِّطَةِ الْحُكُوْمِيَّةِ. هذَا عُمَرُ هُوَ مُتَخَرِّجٌ فِي الْمَدْرَسَةِ الْمُتَوَسِّطَةِ الْمُحَمَّدِيَّةِ. أَنَا فِي الْفَصْلِ الْعَاشِرِ "ب" وَكَذَلِكَ عُمَرُ. ذَلِكَ مُحَمَّدٌ، هُوَ تِلْمِيذٌ فِي الْفَصْلِ الْعَاشِرِ " أ ". هُوَ تِلْمِيذٌ مِثَالِي وَمَاهِرٌ فِي اللُّغَةِ الْعَرَبِيَّةِ، هُوَ يُسَاعِدُنِي فِي تَعَلُّمِ اللُّغَةِ الْعَرَبِيَّةِ.

1. الْمَضْمُونُ الرَّئِيسِيُّ مِنَ النَّصِّ السَّابِقِ هُوَ؟
   أ. يَصِفُ أَحْمَدُ مَدْرَسَتَهُ وَ هِوَايَتَهُ
   ب. يَتَحَدَّثُ أَحْمَدُ عَنْ أُسْرَتِهِ فِي الْمَدْرَسَةِ
   ج. يَتَعَارَفُ أَحْمَدُ نَفْسَهُ وَأَصْدِقَاءَهُ
   د. يُقَدِّمُ أَحْمَدُ عَنْ خِبْرَتِهِ فِي الْمَدْرَسَةِ
   هـ. يُنَاقِشُ أَحْمَدُ الصُّعُوْبَاتِ عَنْ دَرْسِ اللُّغَةِ الْعَرَبِيَّةِ

2. مَاذَا يَقْصِدُ الكَاتِبُ فِي النَّصِ السَّابِقِ؟
   أ. اَلكَاتِبُ يُرَكِّزُ عَلَى وَصْفِ الْمَدْرَسَةِ الثَّانَوِيَةِ الْمُحَمَّدِيَّةِ
   ب. اَلكَاتِبُ يُرَكِّزُ عَلَى عَمَلِيَّةِ التَّعَرُّفِ فِي الْمَدْرَسَةِ
   ج. اَلكَاتِبُ يُرَكِّزُ عَلَى شَرْحِ فَوَائِدِ تَعَلُّمِ اللُّغَةِ الْعَرَبِيَّةِ
   د. اَلكَاتِبُ يُرَكِّزُ عَلَى أَهَمِّيَّةِ الدِّرَاسَةِ فِي الْمَدْرَسَةِ الثَّانَوِيَّةِ
   هـ. اَلكَاتِبُ يُرَكِّزُ عَلَى امْتِيَازِ الدِّرَاسَةِ

3. مَا هُوَ الْمَوْضُوعُ الأَنْسَبُ لِلنَّصِّ السَّابِقِ؟
   أ. اَلصَّدِيْقُ
   ب. اَلدِّرَاسَةُ
   ج. اَلْأُسْرَةُ
   د. اَلتَّعَارُفُ
   هـ. اَلْمُحَادَثَةُ

Conclusion

The development of a HOTS-based Arabic assessment instrument using the *Al-'Ashri* textbook for 10th-grade senior high school students proved to be valid, reliable, and pedagogically meaningful. Statistical analysis indicated that all items met validity standards (r-value > r-table), with a high reliability coefficient (r = 0.88), moderate difficulty levels (0.15–0.85), and acceptable discrimination indices (38.89% good and 61.11% moderate). These findings confirm that the instrument is feasible for evaluating higher-order cognitive abilities aligned with the *Merdeka* Curriculum's objectives to strengthen analytical, evaluative, and creative thinking. Beyond its psychometric soundness, the instrument supports teachers in gradually integrating HOTS-based questions into Arabic learning, combining multiple-choice and open-ended formats to foster reasoning, justification, and creativity. Consequently, HOTS-based assessments can serve not only as evaluation tools but also as instructional strategies that cultivate critical, reflective, and innovative learners in the 21st century.

References

Aflisia, Noza. " تطوير المواد التعليمية النحوية على أساس معرفة المحتوى التربوي التكنولوجي لتنمية مهارات التفكير العليا لدى طلاب قسم تعليم اللغة العربية في جامعة جوروب الإسلامية الحكومية وجامعة فتماواتي سوكرنو الإسلامية الحكومية بنجكولو - Digital Library UIN Sunan Gunung D," 2023. https://digilib.uinsgd.ac.id/71137/.

Aflisia, Noza, Rizka Salsabillah, and Asma Binti Abdul Rahman. "Review of Arabic Language Books Independent Curriculum in Madrasah Ibtidaiyah: Materials, Methods, Advantages, and Disadvantages." *Al-Muktamar As-Sanawi Li Al-Lughah Al-'Arabiyyah (MUSLA)* 2, no. 1 (October 9, 2024): 85–100. https://doi.org/10.24042/albayan.v14i2.12514.

Ainin, Moh. *Penilaian Berpikir Tingkat Tinggi (HOTS) Dalam Pembelajaran Bahasa Arab*. Malang: CV. Bintang Sejahtera, 2020.

Ansari, Bansu Irianto, Muhamad Saleh, Nurhaidah Nurhaidah, and Taufiq Taufiq. "Exploring Students' Learning Strategies and Self-Regulated Learning in Solving Mathematical Higher-Order Thinking Problems." *European Journal of Educational Research* volume-10-, no. volume-10-issue-2-april-2021 (April 2021): 743–56. https://doi.org/10.12973/eu-jer.10.2.743.

Hamidah, Miske Hayunia, and Siti Sri Wulandari. "Pengembangan Instrumen Penilaian Berbasis HOTS Menggunakan Aplikasi 'QUIZIZZ.'" *Efisiensi : Kajian Ilmu Administrasi* 18, no. 1 (May 2021): 105–24. https://doi.org/10.21831/efisiensi.v18i1.36997.

Hikmah, Khizanatul, and Ruli Astuti. "Analisis Perbandingan Kualitas Buku Teks Bahasa Arab Ta'lim Al-Lughoh Al-'Arobiyah Dan Al-'Ashri: Kajian Isi, Penyajian, Dan Bahasa." *Halaqa: Islamic Education Journal* 2, no. 1 (2018): 12–29. https://doi.org/10.21070/halaqa.v1i1.1608.

Madu, Aleksius. "Higher Order Tingking Skills (Hots) In Math Learning." *IOSR Journal of Mathematics (IOSR-JM)* 13, no. 5 (2017): 70–75.

https://doi.org/https://doi.org/10.9790/5728-1305027075.

Mardiyah, Tika, Umi Machmudah, and Abdul Wahab Rosyidi. "Faáliyah Al-Tadribat 'ala Asas Maharat Al-Tafkir Al-Úlya Fi Ta'lim Al-Qiraáh Al-Árabiyyah." *Arabiyatuna : Jurnal Bahasa Arab* 5, no. 2 (October 25, 2021): 289. https://doi.org/10.29240/jba.v5i2.3064.

Primadana, Ramadhan Jabal, Maswani Maswani, Muhbib Abdul Wahab, Achmad Fudhaili, and Raswan Raswan. "Development of Listening and Speaking Skills Teaching Materials Based on CEFR Level." *Arabiyatuna: Jurnal Bahasa Arab* 8, no. 2 (September 30, 2024): 559–78. https://doi.org/10.29240/jba.v8i2.10904.

Qiftiyah, Maratul. "Muatan HOTS Pada Pembelajaran Tematik Materi IPA Kelas 5 Sekolah Dasar." *Scholaria: Jurnal Pendidikan Dan Kebudayaan* 13, no. 1 (January 2023): 28–38. https://doi.org/10.24246/j.js.2023.v13.i1.p28-38.

Ramah, Sutri, and Miftahur Rohman. "Analisis Buku Ajar Bahasa Arab Madrasah Aliyah Kurikulum 2013." *Arabiyatuna : Jurnal Bahasa Arab* 2, no. 2 (2018): 141. https://doi.org/10.29240/jba.v2i2.552.

Sagala, P N, and A Andriani. "Development of Higher-Order Thinking Skills (HOTS) Questions of Probability Theory Subject Based on Bloom's Taxonomy." *Journal of Physics: Conference Series* 1188 (March 2019): 012025. https://doi.org/10.1088/1742-6596/1188/1/012025.

Saputra, Salim, and Tulus Musthofa. "Analisis Buku Teks Pelajaran Bahasa Arab SMA/SMK Muhammadiyah Pada Maharah Al-Kalam Sesuai Standar CEFR." *Lugawiyyat* 4, no. 2 (2022): 81–97. https://doi.org/10.18860/lg.v4i2.17393.

Setiawan, Johan, Ajat Ajat Sudrajat, Aman Aman, and Dyah Kumalasari. "Development of Higher Order Thinking Skill Assessment Instruments in Learning Indonesian History." *International Journal of Evaluation and Research in Education (IJERE)* 10, no. 2 (June 2021): 545. https://doi.org/10.11591/ijere.v10i2.20796.

Widodo, Arif, Zakiyatul Abidah, Nurul Fahmi, and Hana Chebaiki. "Development of Arabic Learning Design Based on Higher Order Thinking Skill with Discovery Learning Model." *Al Mahāra: Jurnal Pendidikan Bahasa Arab* 7, no. 1 (2021): 27–44. https://doi.org/10.14421/almahara.2021.071-02.